

Title: Dataset Analysis Report — Titanic Dataset

Dataset Name: Titanic Dataset

Number of Records: 891 rows and 12 columns

Objective: Predict whether a passenger survived or not.

Data Types

Feature Type	Examples
Numerical	Age, Fare, SibSp
Categorical	Sex, Embarked
Ordinal	Pclass
Binary	Survived

Target Variable

- Survived is the target variable (0 = No, 1 = Yes).

Missing Values

- Age and Cabin contain missing values.
- Age can be filled with mean/median.
- Cabin can be dropped due to high missing percentage.

Class Balance

- Survived has imbalance: more passengers did not survive than survived

ML Suitability

- Dataset size is sufficient for basic machine learning.
- Contains both numerical and categorical data.
- Needs preprocessing (handling missing values, encoding categorical variables).

Conclusion

The Titanic dataset is suitable for machine learning after preprocessing such as handling missing values, encoding categorical features, and balancing classes.